Comparing and Contrasting Data Distributions

Let's investigate variability using data displays and summary statistics.

11.1 Math Talk: Mean

Evaluate the mean of each data set mentally.

- 27 30 33

- 61 71 81 91 101

- 0 100 100 100 100

- 0 5 7 12

11.2 Card Sort: Describing Data Distributions

1. Your teacher will give you a set of cards. Take turns with your partner to match a data display with a written statement.
   a. For each match that you find, explain to your partner how you know it's a match.
   b. For each match that your partner finds, listen carefully to their explanation. If you disagree, discuss your thinking and work to reach an agreement.
2. After matching, determine if the mean or median is more appropriate for describing the center of the data set based on the distribution shape. Discuss your reasoning with your partner. If it is not given, calculate (if possible) or estimate the appropriate measure of center. Be prepared to explain your reasoning.

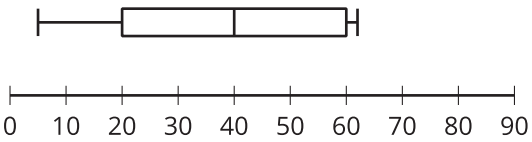
11.3 Visual Variability and Statistics

1. Each box plot summarizes the number of miles driven each day for 30 days in each month. The box plots represent, in order, the months of August, September, October, November, and December.

   a. The five box plots have the same median. Explain why the median is more appropriate for describing the center of the data set than the mean for these distributions.

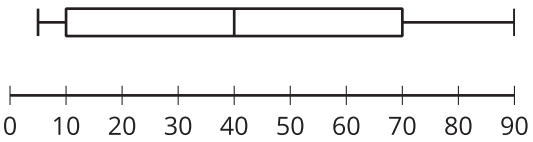
   b. List the box plots in order of least variability to greatest variability. Check with another group to see if they agree.

A



miles driven each day in August

B



miles driven each day in September

C



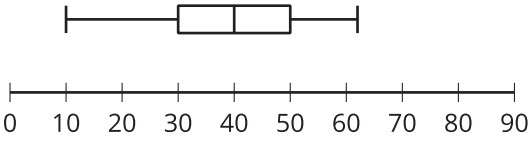
miles driven each day in October

D



miles driven each day in November

E



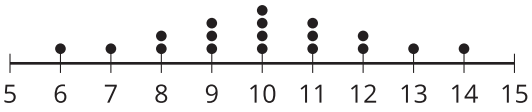
miles driven each day in December

2. These five dot plots represent the number of books on different shelves of the same section of the library.

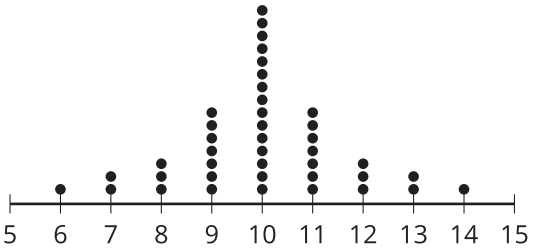
   a. The five dot plots have the same mean. Explain why the mean is more appropriate for describing the center of the data set than the median.

   b. List the dot plots in order of least variability to greatest variability. Check with another group to see if they agree.

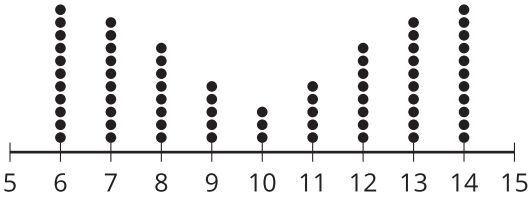
A



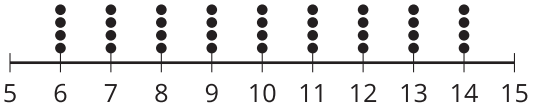
B



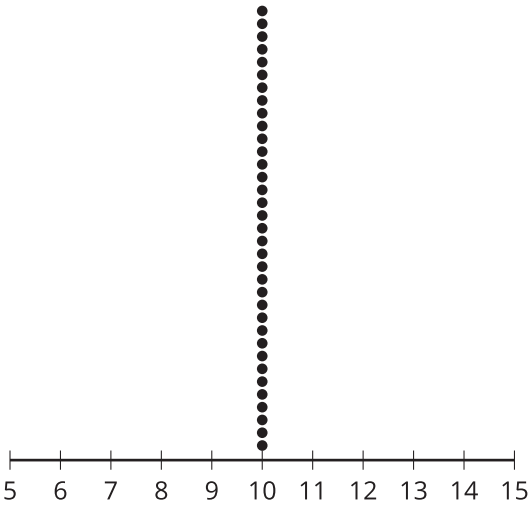
C



D

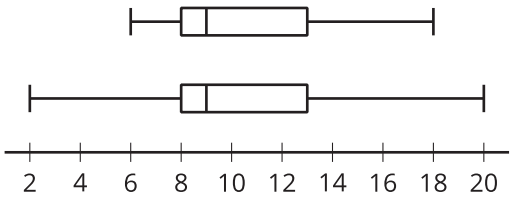


E

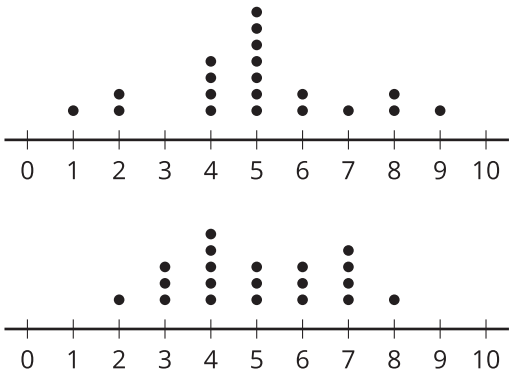


Are you ready for more?

1. These two box plots have the same median and the same IQR. How could we compare the variability of the two distributions?

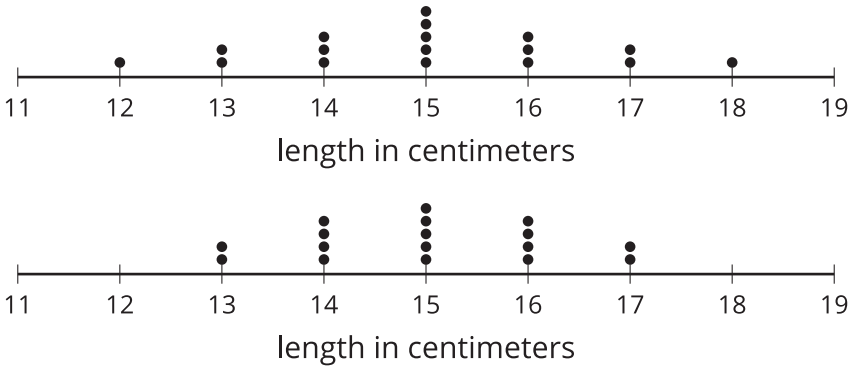


2. These two dot plots have the same mean and the same MAD. How could we compare the variability of the two distributions?



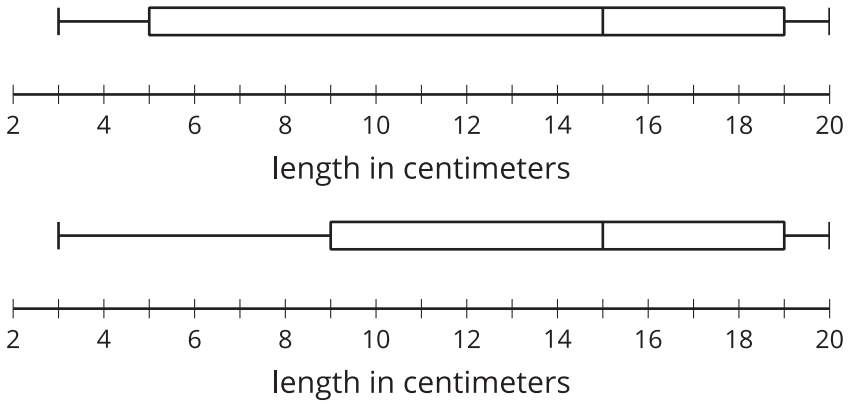
Lesson 11 Summary

The mean absolute deviation, or MAD, is a measure of variability that is calculated by finding the mean distance from the mean of all the data points. Here are two dot plots, each with a mean of 15 centimeters, displaying the length of sea scallop shells in centimeters.



Notice that both dot plots show a symmetric distribution so the mean and the MAD are appropriate choices for describing center and variability. The data in the first dot plot appear to be more spread apart than the data in the second dot plot, so you can say that the first data set appears to have greater variability than does the second data set. This is confirmed by the MAD. The MAD of the first data set is 1.18 centimeters and the MAD of the second data set is approximately 0.94 cm. This means that the values in the first data set are, on average, about 1.18 cm away from the mean, and the values in the second data set are, on average, about 0.94 cm away from the mean. The greater the MAD of the data, the greater the variability of the data.

The interquartile range, IQR, is a measure of variability that is calculated by subtracting the value for the first quartile, Q1, from the value for the third quartile, Q3. These two box plots represent the distributions of the lengths in centimeters of a different group of sea scallop shells, each with a median of 15 centimeters.



Notice that neither of the box plots have a symmetric distribution. The median and the IQR are appropriate choices for describing center and variability for these data sets. The middle half of the data displayed in the first box plot appear to be more spread apart, or show greater variability, than the middle half of the data displayed in the second box plot. The IQR of the first distribution is 14 cm, and the IQR is 10 cm for the second data set. The IQR measures the difference between the median of the second half of the data, Q3, and the median of the first half, Q1, of the data, so it is not affected by the minimum or the maximum value in the data set. It is a measure of the spread of the middle 50% of the data.

The MAD is calculated using every value in the data set, and the IQR is calculated using only the values for Q1 and Q3.